Elektrotehniški vestnik 60(4): 240–250, 1993
Electrotechnical Review, Ljubljana, Slovenija

A Direct Recovery of Superquadrics in Range Images Using Recover-and-Select Paradigm

Aleš Leonardis, Franc Solina, Alenka Macerl

Computer Vision Laboratory, Faculty of Electrical Engineering and Computer Science, University of Ljubljana, Tržaška 25, 61001 Ljubljana, Slovenia

E-mail: Ales.Leonardis@fer.uni-lj.si, Franc.Solina@fer.uni-lj.si

Abstract. In this article we present a novel approach to reliable and efficient recovery of part-descriptions in terms of superquadric models. In contrast to common beliefs that the recovery of volumetric models is only possible after the data has been pre-segmented, usually in a hierarchical fashion using local surface properties, curvature etc., we show that a set of superquadric models can be *directly* recovered from unsegmented range images. This is achieved in the *recover-and-select* paradigm which consists of two intertwined stages: model-recovery and model-selection. At the model-recovery stage a redundant set of superquadrics is initiated in the image and allowed to grow, which involves an iterative procedure combining data classification and parameter estimation. All the recovered models are passed to the model-selection procedure, which is defined as a quadratic Boolean problem, where only the models resulting in the simplest overall description are selected. By combining model-recovery and model-selection in an iterative scheme a computationally efficient procedure is achieved. We present experimental results on several real range images.

Key words: Part-level range image segmentation, direct recovery of volumetric models, superquadrics in recover-and-select paradigm.

Neposredna rekonstrukcija super-elipsoidov na globinskih slikah s pomočjo paradigme gradnje in izbire modelov

Povzetek. V članku predstavljamo zanesljivo in učinkovito metodo za gradnjo opisov globinskih slik z modeli super-elipsoidov. Super-elipsoidi so se izkazali kot primerni prostorski modeli za opis naravnih in tehničnih predmetov po delih. V nasprotju s splošnim prepričanjem, da je rekonstrukcija prostorskih modelov možna le na predhodno segmentiranih podatkih, kar običajno vključuje metode diferencialne geometrije, pa v tem članku pokažemo, da je možno modele super-elipsoidov zgraditi neposredno na danih podatkih. To dosežemo s pomočjo *paradigme gradnje in izbire modelov*, ki jo sestavljata dva postopka: gradnja modelov in izbira modelov. Prvi postopek temelji na sočasni gradnji redundančne množice modelov z iterativno metodo, ki združuje ugotavljanje pripadnosti slikovnih elementov posameznim modelom in ocenjevanje parametrov modelov, ki jim ti elementi pripadajo. Drugi postopek, definiran kot kvadratični Boolov problem, pa izmed vseh zgrajenih modelov, ki predstavljajo možne kandidate za končni opis slike, izbere tiste, ki opišejo slikovne podatke na najenostavnejši način. Z iterativo povezavo postopkov za gradnjo modelov in njihovo izbiro dosežemo potrebno računsko učinkovitost. V članku so podani eksperimentalni rezultati, ki smo jih dobili na vrsti globinskih slik.

Ključne besede: Segmentacija globinskih slik, neposredna gradnja prostorskih modelov, super-elipsoidi v paradigmi gradnje in izbire modelov.

1 Introduction and Motivation

The significance of detecting geometric structures in images has long been realized in the vision community. One of the primary intentions has been to build primitives that would bridge the gap between low-level features and high-level symbolic structures useful for further processing. We advocate the view that the purpose of machine vision is not to reconstruct the scene in its entirety, but rather to search for specific features that enter, via data aggregation (i.e. model recovery), into a symbolic description of the scene necessary to achieve a specific task.

Many theories have emerged which emphasize the importance of extraction of perceptually relevant image structures [27,26,9,30,29] showing that a model-free interpretation is doomed to fail due to the underconstrained nature of the problem. The relevant image structures essentially encode the knowledge or expectations of how the data is structured and help augmenting imperfect vi-

Received 23 September 1993
Accepted 22 October 1993

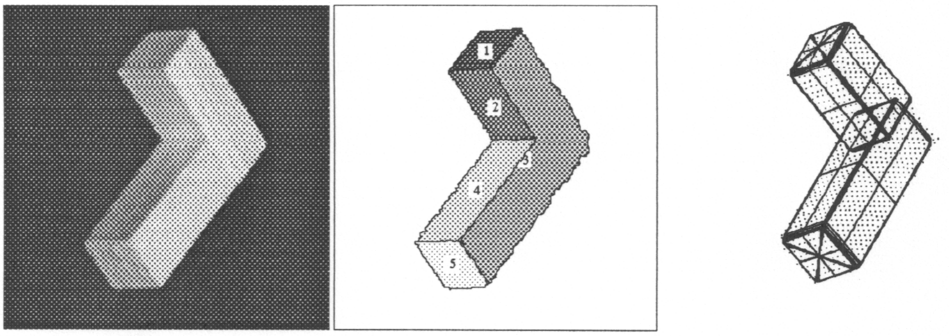


Figure 1. An L-shaped object. Planes of the surface-level description can not be simply joined into volumetric primitives.

sual data with intrinsic information, thus enabling and making the recovery process more robust.

To represent the "natural" structuring of the world and support recognition and learning of such "natural" structures from images, people employ a part structure. Perceptually, the world can be broken down into parts, and the goal of computer vision is to recover from images this part structure (segmentation) and the metric properties of individual parts (shape recovery). Two types of volumetric models, generalized cylinders [10,27,29] and superquadrics [5,30,36] have emerged for such part-level modeling. Lately, superquadric models are gaining popularity in the vision and robotics community because of its compact representation and robust recovery methods of individual models. The constraint that the models must be bounded volumes has been used, for example, for robot grasping [2,16], for environment modeling in robot path-planning [12], or for modeling kinematic chains [1,39].

The motivation for this work is that volumetric part models have been successfully used only for shape recovery of individual parts but not for segmentation of images into parts. This is due to the difficulty of simultaneous classification (grouping) of image elements and of model parameter estimation which has been a major obstacle to successful applications that require reliable extraction of volumetric models from the data.

Although rigorous schemes for recovery of volumetric models have been developed, most of them make the assumption that the segmentation problem has been solved by some other means [7,20,36,34,38]. By trying to avoid, explicitly or implicitly, the phase of classifying and grouping together the image elements that form individual volumetric models, the real complexity of the task is obscured. We agree that certain insights can be gained by observing problems separately, but we believe that in this case it completely obscures the difficulty of extracting complex parametric descriptions from real images. While searching for rigid models in an image can be performed on a global scale, parametric models have to be initiated in a local neighborhood in order to prevent data points from different descriptions to be confounded together.

There have been several attempts to segment and recover volumetric models from the data. These approaches usually involve several procedures, mostly applied in a hierarchical fashion, ranging from the estimation of local surface properties, curvature, etc. to more complex, such as symmetry seeking, in order to partition the data into parts that can supposedly be represented with a single volumetric model. These approaches, in fact, isolate segmentation stage from the representation stage and significant efforts are necessary to combine, usually surface-type descriptions into volumetric models. The ability to even identify a set of surfaces as belonging to a given volume is not a trivial task without knowing at least connectedness, and preferably surface closure. Moreover, a surface-level description may not be consistent with the volumetric description. Figure 1 shows an L-shaped object* whose volumetric description can not be obtained by a simple combination of recovered surfaces.

In this article we demonstrate that a specific set of volumetric models, i.e. superquadrics can be *directly* recovered from range data, which is in contrast to common beliefs that the recovery of volumetric models is possible only after the data has been pre-segmented using extensive pre-processing. To achieve this goal we have cast the problem of volumetric recovery [36] in the *recover-and-select* paradigm for the recovery of geometric parametric structures from image data [25]. We have been motivated for this new application of the recover-and-select paradigm by the encouraging results that the paradigm achieved on a variety of domains: segmentation and recovery of surface models in range data [22,24] and curve-models [23] in intensity images.

The article is organized as follows: in section 2 we discuss the related work. In section 3 we explain the choice of geometric volumetric primitives, namely *superquadrics*. An outline of the *recover-and-select* paradigm is given in section 4. In section 5 we describe specific details that pertain to the recovery of superquadric models. Some of our experimental results are shown in section 6. In conclusion we summarize our paradigm and outline the work in progress.

*All range images in this article except the one in Fig. 5 were kindly provided by Marjan Trobina from ETH, Zürich, Switzerland.

2 Background and related work

Segmentation of images into regions corresponding to single objects or their parts is one of the harder problems in computer vision. There were several attempts to define parts as perceived by human vision in mathematical terms. Koenderink and van Doorn [19] defined part intersections as parabolic lines on the surfaces of objects. Hoffman and Richards [17] refined the solution by using instead the "negative minima of principal curvatures". The application of the above partitioning rules on real images is difficult because of imperfections in low-level shape descriptions.

Most approaches to segmentation in computer vision are based on using local image information, in the form of low level image models such as edges, surface patches and surface normals. The resulting segmentation is often arbitrary, because merging or growing of such small surface regions essentially still relies on local information. If such local segmentation methods are made sensitive enough to detect subtle changes in first or second derivatives in order to find part boundaries, they become susceptible to noise and details that are not relevant for the targeted level of representation. Essentially, local pieces of information cannot decide on the shape of the whole part if the concept of the whole part is not well defined as such.

The problem of using part boundaries to define the shape of parts can be circumvented by defining the whole part shape directly. Biederman [9] argued for a set of primitive building blocks which can describe the wealth of different shapes by combining them like phonemes in a language. Biederman proposed a set of 35 primitives, called Geons, which he obtained by analyzing non-accidental changes on a generalized cylinder.

Pentland [30] proposed the use of superquadric models combined with global deformations as a set of primitives which very closely correspond to Hoffman's notion of parts [17] and which could be recovered directly from images. Pentland's initial idea to analytically solve all independent superquadric parameters did not prove to be practical. Pentland [31] later combined part model recovery with segmentation and based it on a coarse search through the entire superquadric parameter space. A goodness-of-fit function was evaluated at selected points in the parameter space for many largely overlapping range image regions. This segmentation stage was later followed by a gradient-descent optimization for individual superquadric models. However, this method is computationally very expensive and the segmentation results still not precise.

Solina and Bajcsy [3,36] formulated the recovery of deformed superquadric models from range data as a least-squares minimization of a fitting function. They have approached the problem of segmentation by a recursive splitting technique that starts fitting a single superquadric model to the whole scene, shrinks the model until it fits one of the parts and starts the same process with a new model on the remainder of the scene [3]. While they solved the single model recovery problem successfully, the question of efficient and reliable segmentation remained open.

Pentland [33] proposed another segmentation method based on 2-D silhouettes (2-D projections of 3-D superquadric parts of different shapes and of different orientations). He devised a robust method for finding such 2-D silhouettes at different scales in the image. Once part segmentation was accomplished 3-D superquadric models were fitted to individual part regions. Superquadric fitting was based on numerical minimization using as the error metric the squared distance along the depth axis z between the range data and the projected volume's visible surface. In related work on recovery of superquadric-like physically based models Pentland presegmented range images using simple polynomial shape models [34].

Ferrie *et al.* [13] proposed a hierarchical scheme which at the highest level achieves segmentation in terms of superquadric part models. The approach consists of several stages performed in a purely bottom-up strategy. These stages are: darboux frames, snakes, and superquadrics. There are several major drawbacks that can be identified with this approach. Since the data points are grouped together on the basis of differential geometry regardless of the final parametric models, thus essentially decoupling the classification and representation phases, the resulting segments might not correspond to the final primitives.

Gupta and Bajcsy [15] based the recovery of superquadrics on the result of surface segmentation [22]. Region (surface) adjacency information, surface discontinuities, and global shape properties are used to guide the volumetric segmentation. However, situations like the one illustrated in Fig. 1 make this approach excessively complicated.

We think that in all of the above approaches the predictive power of volumetric models is not used to its full potential since mostly rules for combining low level models into larger ones are used. We believe that when higher level generic models are well defined as in the case of superquadrics, one can attempt to find them in a more direct way.

It is important not only to use the information that is inherent in the choice of parametric models, but also to use it efficiently to assure the consistency and coherence of procedures throughout the segmentation/shape recovery stages. In other words, it is not only desirable, but also essential for an efficient and reliable recovery of models to employ the representation to guide detection and grouping processes [8,4].

3 Parametric models

The criteria for the selection of geometric primitives have been studied extensively by vision researchers (see [11] and the references therein). It is generally accepted that the primitives should be

- invariant to affine transformations of the data set,
- stable to minor changes of the data due to noise or viewpoint,

- accessible.

The last requirement, *accessibility*, defined as computability of the primitive, is essential since our goal is to recover a structure from an input. For example, the primitives should have local support, so that they can cope with occlusions. This requirement not only constrains the choice of the primitives but imposes certain conditions on the model-recovery procedure as well.

In addition, primitives should balance the trade-off between data reduction and faithfulness to measured data. In terms of parametric models, this means the number of parameters needed to describe a unit of a spatial extent.

Superquadrics are models that satisfy the above mentioned criteria. However, all model based approaches are restricted, since they cannot model everything present in the input data. But what is important, is that the method signals when the models are inadequate to describe the data, so that a different type of models can be invoked.

3.1 Superquadric models

Superquadrics are an extension of basic quadric surfaces and solids. They have been considered as volumetric primitives for shape representation in computer graphics [5] and computer vision [30,36,38,28,13]. The reason being that they are convenient part-level models that can be further deformed and glued together to model articulated objects.

Superquadric surface is defined by the following equation

$$F(x,y,z) = \left(\left(\left(\frac{x}{a_1}\right)^{\frac{2}{\epsilon_2}} + \left(\frac{y}{a_2}\right)^{\frac{2}{\epsilon_2}} \right)^{\frac{\epsilon_2}{\epsilon_1}} + \left(\frac{z}{a_3}\right)^{\frac{2}{\epsilon_1}} \right). \tag{1}$$

When both exponents ϵ_1 and ϵ_2 equal 1, the surface forms an ellipsoid or, if a_1, a_2, a_3 are all equal, a sphere. When $\epsilon_1 \ll 1$ and $\epsilon_2 = 1$, the superquadric surface is shaped like a cylinder. Parallelepipeds are produced when both $\epsilon_1 \ll 1$ and $\epsilon_2 \ll 1$. Modeling capabilities of superquadrics can be enhanced by deforming them in different ways, such as global tapering and bending [6,36] or local deformations for detailed modeling [38]. In the next subsection we will summarize a robust method for recovery of isolated superquadric models, based on a nonlinear least-squares method [36].

3.2 Recovery of superquadrics

The implicit function (Eq. 1) defined in an object centered coordinate system determines where a given point (x, y, z) lies relative to the superquadric surface. To recover a superquadric in a general position, the implicit function for general position is used

$$F(x,y,z) = F(x,y,z; a_1, a_2, a_3, \epsilon_1, \epsilon_2, \phi, \theta, \psi, p_x, p_y, p_z). \tag{2}$$

This expanded "inside-outside" function has 11 parameters; a_1, a_2, a_3 define the superquadric size; ϵ_1 and ϵ_2 are shape parameters; ϕ, θ, ψ define the orientation in space, and p_x, p_y, p_z define the position in space. We refer to the set of all model parameters as $\boldsymbol{\Lambda} = \{a_1, a_2, \ldots, a_{11}\}$.

Suppose we have N 3-D points on a surface of an object (x_W, y_W, z_W) which we want to model with a superquadric. We want to vary the 11 parameters $a_j, j = 1, \ldots, 11$ in equation (2) to get such values for a_j's that most of the 3-D points will lie on, or close to the model's surface. There will probably not exist a set of parameters $\boldsymbol{\Lambda}$ that perfectly fits the data. Finding the model $\boldsymbol{\Lambda}$ for which the distance from points to the model's surface is minimal is a least-squares minimization problem. Since due to self occlusion, not all sides of an object are visible at the same time, we introduced an additional constraint. Among all possible solutions we search for the *smallest* superquadric that fits the given range points in the least squares sense. We defined the following function which has a minimum corresponding to the smallest superquadric that fits a set of 3-D points

$$f = \sqrt{a_1 a_2 a_3}\,(F^{\epsilon_1} - 1). \tag{3}$$

Since, for a point (x_W, y_W, z_W) on the surface of a superquadric

$$f(x_W, y_W, z_W; a_1, \ldots, a_{11}) = 0, \tag{4}$$

we have to find

$$G = min \sum_{i=1}^{N} [f(x_{W_i}, y_{W_i}, z_{W_i}; a_1, \ldots, a_{11})]^2. \tag{5}$$

Since f is a nonlinear function of 11 parameters $a_j, j = 1, \ldots, 11$, minimization must proceed iteratively. Given a trial set of values of model parameters $\boldsymbol{\Lambda}_k$, we evaluate equation (3) for all N points and employ a procedure to improve the trial solution. The procedure is then repeated with a set of new trial values $\boldsymbol{\Lambda}_{k+1}$ until the sum of least squares (5) stops decreasing, or the changes are statistically insignificant. In most cases 15 iterations are more then sufficient. We use the Levenberg-Marquardt method for nonlinear least squares minimization since first derivatives $\delta f/\delta a_i$ for $i = 1, \ldots, 11$ can be computed analytically. We raised the "inside-outside" function F (Eq. 1) to the power of ϵ_1 in Eq. (3) to make the error metric more quadratic and more suited for rapid convergence.

It turns out that only very rough initial estimates of object's true position, orientation, and size suffice to assure convergence to a local minimum that corresponds to the actual shape. Initial values for both shape parameters, ϵ_1 and ϵ_2 are set to 1, which means that the initial model $\boldsymbol{\Lambda}_E$ is always an ellipsoid. Position in world coordinates is estimated by computing the center of gravity of all range points, and the orientation is estimated by computing the central moments with respect to the center of gravity. The initial model $\boldsymbol{\Lambda}_E$ is oriented so that the axis z of the object centered coordinate system lies along the longest side (axis of least inertia) of the object. Global deformations of superquadrics require additional parameters that can be recovered in the same way [36], but in this article we are going to use only the non-deformed shapes.

The fitting function (3) can be regarded as an energy function on the space of model parameters. Minimization methods can, in general, only guarantee convergence to a local minimum. The starting position in the parameter space ($\boldsymbol{\Lambda}_E$) determines to which minimum will the minimization procedure converge. To assure that the minimization procedure does not get stuck in a shallow local minimum, we add noise during the minimization procedure.

This recovery method which is described in detail in [36] and the herein proposed error metric has been used by several other researchers [2,15,16,35,13]. Although the distance metric (Eq. 3) varies across the surface when $\epsilon_1 \neq \epsilon_2$ and when size parameters change it has been proven to be efficient and robust.

In the next section we will first summarize the *recover-and-select* paradigm and then describe the modifications necessary for the inclusion of superquadric models into it.

4 Recover and select

We present here a general outline of the recover-and-select paradigm. The choice of specific models, in our case superquadrics, imposes certain constraints on the recovery of these structures from images, which will be described in section 5.

4.1 Model recovery

Recovery of parametric models is difficult because one has to solve **two** problems:

1. find image elements that belong to a single parametric model,
2. *and* determine the values of the parameters of the model.

For image elements that have already been classified (segmented) one can determine the parameters of a model by applying standard statistical estimation techniques. Conversely, knowing the parameters of the model, a search for compatible image points can be accomplished by pattern classification methods. Thus we propose to solve these **two** problems simultaneously by an iterative method, conceptually similar to the one described by Besl [8], which combines data classification and model fitting.

One of the crucial dilemmas is where to find the initial estimates (seeds) in an image since their selection has a major effect on the success or failure of the overall procedure. We propose that a search for the points that could belong to a single parametric model is performed in a grid-like pattern of windows overlaid on the image. Thus, the requirement of classifying all data points of a certain model is relaxed to finding only a small subset. However, there is no guarantee that every seed will lead to a good description since some initial models can be constructed over areas which belong to different models. As a remedy we propose to *independently* build *all possible* models using all statistically consistent seeds and to use them as hypotheses that could compose the final description.

Having an initial set of points (a seed) we estimate the parameters of the model. If sufficient similarity between the model and the data is established, ultimately depending on the task at hand, we proceed with a search for more compatible points. An efficient search is performed in the vicinity of the present border points of the model. New image elements are included in the data set and the parameters of the model are updated. The new goodness-of-fit is computed and compared to the old value. This is followed by a decision whether to perform another iteration or terminate the procedure.

The main features of model-recovery procedure are: a high degree of resistance to outliers since the performance of the fitting is constantly monitored and an independent and potentially parallel execution of the recovery procedure for individual models.

The final outcome of the model-recovery procedure for a particular model M_i consists of three terms which are subsequently passed to the model-selection procedure: the set of data elements n_i that belong to the model M_i, the type of the parametric model and the corresponding set of parameters of the model (N_i denotes the cardinality of this set), and the goodness-of-fit value ξ_i which describes the conformity between the data and the model.

4.2 Model selection

The redundant representation obtained by the model-recovery procedure is a direct consequence of the decision that a search for parametric volumetric models is initiated everywhere in an image. Several of the models are completely or partially overlapped. The task of combining different models is reduced to a selection procedure which considers many competitive solutions and selects those that produce the simplest description, i.e., the one that accounts for the largest number of data points while keeping the deviations between data points and models low. Intuitively, this reduction in complexity of a representation coincides with a general notion of simplicity which has a long history in psychology (Gestalt principles). The formalization of this principle led in information theory to the method of *Minimum Description Length* MDL, which has recently found its way to computer science, including computer vision [21,14,32].

The objective function $F(\mathbf{m})$, which is to be maximized in order to produce the "best" description in terms of models, has the following form:

$$F(\mathbf{m}) = \mathbf{m}^T \mathbf{Q} \mathbf{m} = \mathbf{m}^T \begin{bmatrix} c_{11} & \dots & c_{1M} \\ \vdots & & \vdots \\ c_{M1} & \dots & c_{MM} \end{bmatrix} \mathbf{m} \ , \tag{6}$$

where $\mathbf{m}^T = [m_1, m_2, \dots, m_M]$. m_i is a *presence variable* having the value 1 for the presence and 0 for the absence of the model M_i in the final description. The diagonal terms of the matrix $\mathbf{Q}$ express the cost-benefit value of a particular model M_i,

$$c_{ii} = K_1 n_i - K_2 \xi_i - K_3 N_i \ . \tag{7}$$

K_1, K_2, K_3 are weights which can be determined on a purely information-theoretical basis (in terms of bits), or they can be adjusted in order to take into account the signal-to-noise ratio.
The off-diagonal terms handle the interaction between the overlapping models

$$c_{ij} = \frac{-K_1 \Gamma(M_i, M_j) + K_2 \xi_{i,j}}{2} \ . \tag{8}$$

$\Gamma(M_i, M_j) = |M_i \cap M_j|$ is the number of points that are explained by both models. $\xi_{i,j}$ corrects the diagonal error terms in case that both models are selected

$$\xi_{i,j} = \max\Big(\sum_{\Gamma(M_i,M_j)} d^2_{M_i}, \sum_{\Gamma(M_i,M_j)} d^2_{M_j} \Big) \ . \tag{9}$$

The error terms $d^2_{M_i}$ and $d^2_{M_j}$ are calculated in the area of intersection and correspond to deviations from the i-th and j-th model, respectively. Notice that in the intersection area where both models cover the data the smaller error is taken.

Maximizing the objective function $F(\mathbf{m})$ belongs to the class of problems known as combinatorial optimization (Quadratic Boolean problem). Since the number of possible solutions increases exponentially with the size of the problem, it is usually not tractable to explore them exhaustively. The exact solution has to be sacrificed to obtain a realizable solution.

Various methods were proposed for finding a "global extreme" of a class of non-linear objective functions that were previously thought to be practically insoluble. However, they are in general too time consuming to be applicable in most situations. It turns out that in our case, due to the specific nature of the problem, a reasonable solution can be obtained by a direct application of the *greedy algorithm* which at any individual stage selects the option which is locally optimal. In other words, the models are selected in the sequence that corresponds to the size of their contributions to the objective function, which is equivalent to applying at each stage of the algorithm the *winner takes all* principle. This is a simple mechanism that can account for a number of phenomena that take place in brains [18].

The algorithm is simple: We start with the state in which no models have been selected. The initial value of $F(\mathbf{m})$ is 0. A successor state is formed by adding a not yet selected model to the current description. The selected model is the one which contributes the most to the value of the objective function. The process of adding models to the current description is repeated as long as the value of the objective function can be increased.

A discussion on when the greedy algorithm produces satisfactory results, and the computational complexity of the method is given in [25].

4.3 Model recovery and model selection

In order to achieve a computationally efficient procedure the model-recovery and model-selection procedures are combined in an iterative fashion. The recovery of currently active models is interrupted by the model-selection procedure which selects a set of currently optimal models which are then passed back to the model-recovery procedure. This process is repeated until the remaining models are completely recovered. The trade-offs which are involved in the dynamic combination of these two procedures are discussed elsewhere [25].

5 Superquadrics in the recover-and-select paradigm

In this section we give some details of the superquadric model recovery as used in the recover-and-select paradigm. The overall processing scheme is given in Table 1. Seed selection, decision making (when to expand or stop growing superquadric models), and search for new compatible points (growing) are given in the following subsections.

Table 1. Algorithm for superquadric recovery in the *Recover-and-Select* paradigm.

input: a range image

determine a set of seeds

for all seeds **do**
    fit a SQ model (estimate parameters)
    if (GOF == OK) put the SQ model into the set of
        currently *active, not fully grown* models
    else the seed is rejected

while there are any *active, not fully grown* SQ models **do**
    for all *active not fully grown* SQ models **do**
        extrapolate/find compatible points
        if no new compatible points
            the SQ model is *fully grown*
        else
            fit a SQ model
            if (GOF $\neq$ OK)
                reject the last included points
                the SQ model is *fully grown*
            end if

        end if
    end for

    perform selection among all *active* SQ models
    for the current optimal description
    (only the selected SQ models remain *active*)

end do

output: part-level (SQ) description of a range image

5.1 Seed selection

Initial seeds are placed on the image in the grid-like pattern of windows. An initial seed encompasses a set of range data points in a small window whose size is determined on the basis of scale and can be adaptively changed depending on the task. A superquadric model is fitted to the data set. Next, a decision is made whether all the data points belong to that single model. This decision, which is based on the goodness-of-fit measure (see the next subsection on *decision making*), is not critical due to the redundant nature of the paradigm. It eliminates those seeds that were placed on data sets that cross part boundaries and helps reducing the number of seeds at start of processing.

5.2 Decision making

A decision whether a model should grow further or not, depends on the established similarity between the model and the data. If sufficient similarity is established, ultimately depending on the task at hand, we accept the currently estimated parameters, together with the current data set, and proceed with the search for more compatible points. The question is what could be used as a goodness-of-fit measure. Due to its dependence on the superquadric size and shape parameters $(a_1, a_2, a_3, \epsilon_1, \epsilon_2)$ the algebraic distance (inside-outside function – Eq. 1) is not suitable. It turns out that an approximation to the Euclidean distance does the job when the distance of the point from the corresponding superquadric model is small. In fact, due to the strategy of acquiring new data points based on the consistency with the current model, this condition is always satisfied. The approximate Euclidean distance for a given data point is given by:

$$d^2((x,y,z), \mathbf{\Lambda}) \approx \frac{f^2(x,y,z)}{\| \nabla f((x,y,z)) \|^2}\,, \qquad (10)$$

where f is given in equation (3). The sum over all data points belonging to the model determines the goodness-of-fit of the entire model. This is also the measure which is together with the number of data points encompassed by the model passed to the selection procedure.

5.3 Search for new compatible points

In accordance with the paradigm, an efficient search for more compatible points is performed in the vicinity of the present border points of the model. This is achieved first by simply multiplying the size of the current model (a_1, a_2, a_3) by factor of 2 to get an enlarged model. Then all points which are inside this enlarged model are checked on how close are they to the surface of the original model (again, the approximate Euclidean distance was used (Eq. 10)). Only those points that are close enough to the original model are included in the updated set of points. On this set of points a new superquadric model-recovery procedure is started.

The fitting function f used in superquadric model recovery was skewed so that points on the outside of the superquadric model have a smaller influence on its current shape. Points on the inside of the model produce a larger error than those on the outside. In this way, superquadric models in the final set can touch each other and the points belonging to each model have no influence on other models.

6 Experimental results

The proposed method has been tested on a variety of range images. All the examples were run on workstation HP-715/50. Original range images were subsampled to speedup the recovery. This probably caused the slight misalignements of final models in Figs. 1 and 2. Processing of examples shown took on the average less than 10 minutes. However, processing time is not critical since individual models could be recovered in parallel. Besides, the computation of the superquadric fitting function and its derivatives is independent for each range point and can also be parallelized in a straightforward way.

The thresholds in the model-recovery procedure (compatibility threshold, acceptable goodness-of-fit) were determined on the basis of the analysis of the noise properties of the acquisition process.

Four examples of processed range images are shown. Each figure shows the following image sequence:

- **(a)** the intensity image of the object,
- **(b)** the corresponding range image,
- **(c)** the original image resampled and transformed into the form (x,y,z), appropriate for the superquadrics recovery process,
- **(d)** the recovered volumetric models after the first model selection,
- **(e)** volumetric models at the midpoint of the recover-and-select process,
- **(f)** final result (superquadric models on top of only those range points that influenced the models).

The results are commented in the captions of each figure.

7 Conclusions

We have successfully combined the two existing methods, namely recovery of superquadric models [36] and the *recover-and-select* paradigm [25]. Thus we showed that a direct segmentation into part-level volumetric models is possible. The interpretation of these models is straightforward, with direct applications for manipulation, object recognition and CAD modeling (reverse engineering).

There are several open issues that we are going to address in our future work. An immediate problem is merging of models that represent different sides of the same physical part separated by occlusion (see Fig. 4).

We also plan to explore the possibility to use the proposed method for recovering (in parallel) multiple geometric modalities (surfaces and volumes) and different image modalities (range and intensity images). The final description of a scene would result from a selection

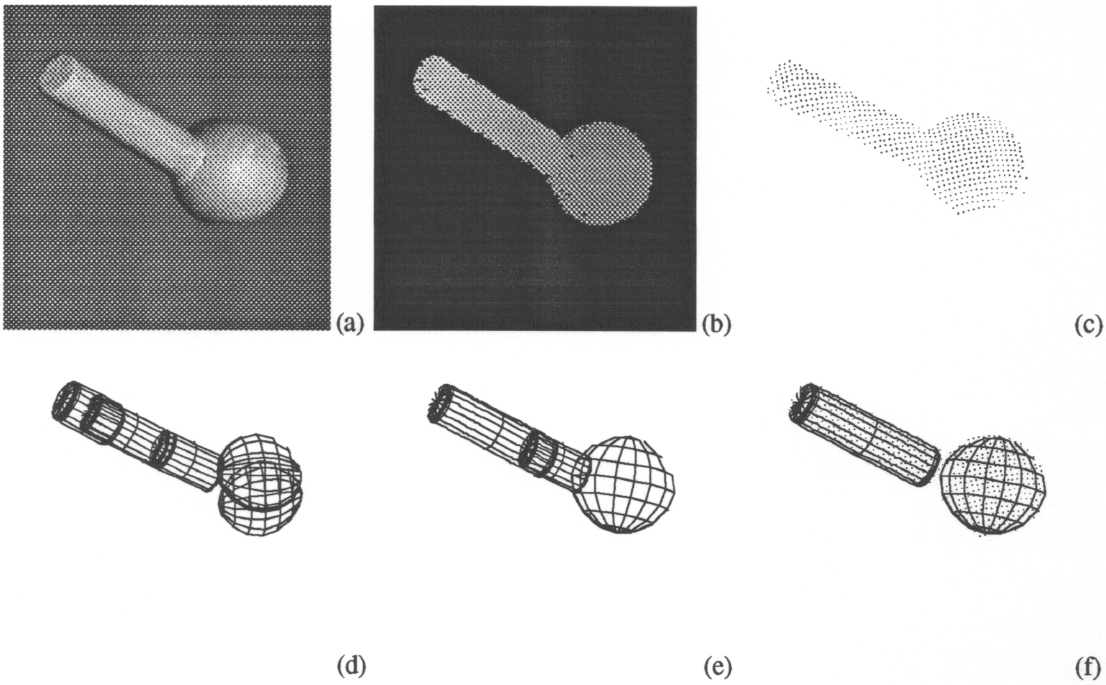


Figure 2. A sphere atop of a cylinder. Although 25 models were instantiated at the start on this range image the recover-and-select procedure selects for the final representation only two fully grown superquadric models.

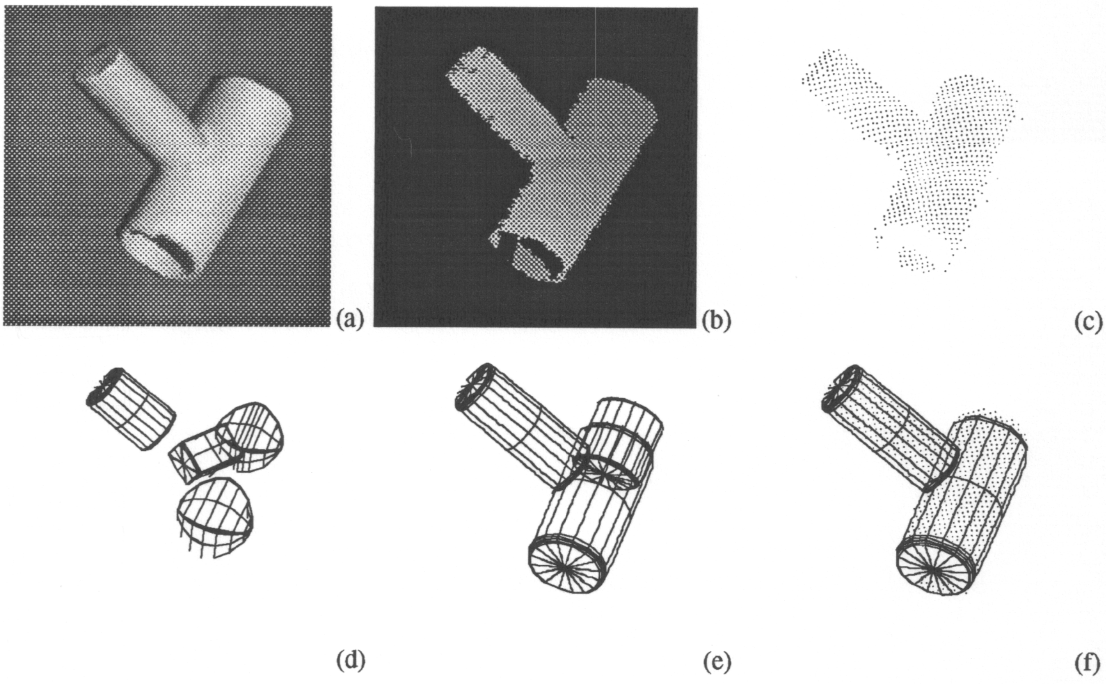
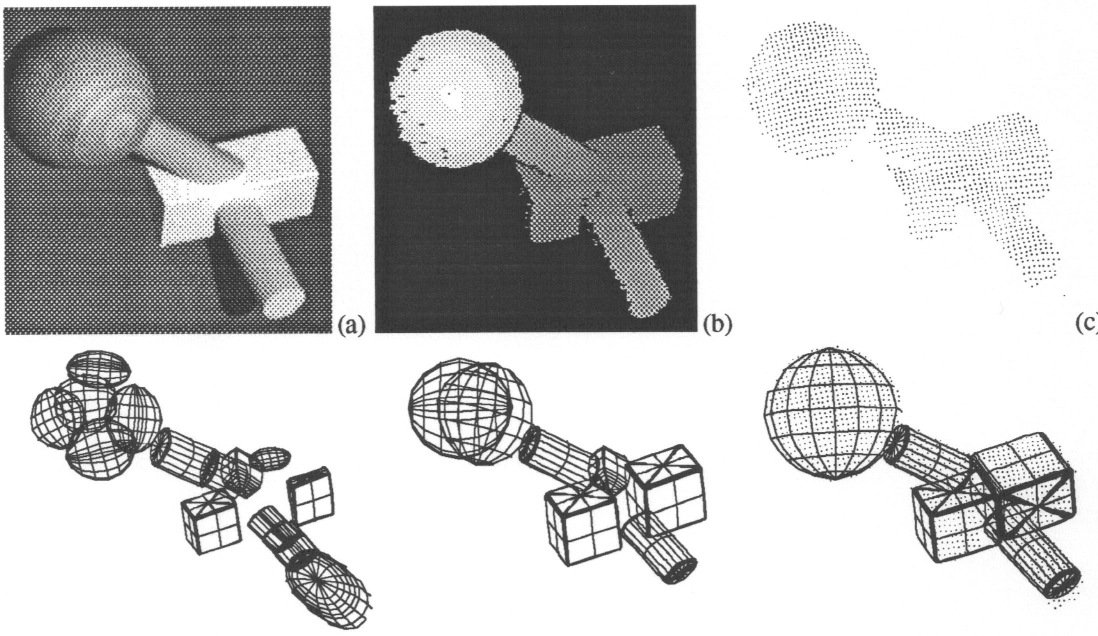


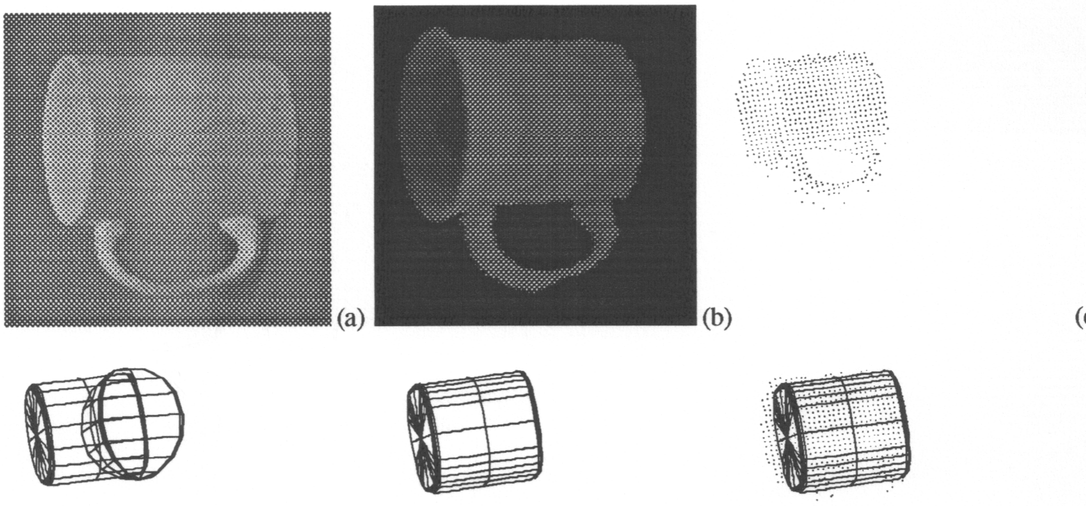
Figure 3. A tube and a cylinder that intersect. Twenty seeds were placed on the range image at the start of processing. The final representation is most compact, consisting of two superquadrics. The points on the inside of the tube are also incorporated into the final description (f).



(a) (b) (c)

(d) (e) (f)

Figure 4. A complex scene consisting of a sphere and two cylinders that are joined with a block. Forty seeds were placed on the scene at start. The final description (f), however, consists of a sphere, two cylinders and *two* blocks. Neither of the two superquadric models in figure (e) could grow into the other side of the block because the number of range points that bridge the gap caused by occlusion is not sufficient.



(a) (b) (c)

(d) (e) (f)

Figure 5. This range image of a cup illustrates that only those parts of the image that can be modeled with the chosen set of models are in fact included into them. Hence range points on the handle of the cup are missing from the final result (f). Modeling of the handle would require a deformed superquadric or some surface models.

procedure (employing the MDL principle), as proposed in [37].

Acknowledgement

This research was supported in part by The Ministry for Science and Technology of The Republic of Slovenia (Project P2-1122).

8 References

[1] E. I. Agba, T. Wong and M. Z. Huang. Solid modeling of kinematic chains. *Fifth International conference on advanced robotics, Robots in unstructured environments*, pages 959–963, Pisa, Italy, 1991. IEEE.

[2] P. K. Allen and P. Michelman. Acquisition and interpretation of 3-D sensor data from touch. *IEEE Transactions on Robotics and Automation*, 6(4):397–404, 1990.

[3] R. Bajcsy and F. Solina. Three dimensional shape representation revisited. In *Proceedings First International Computer Vision Conference*, pages 231–241, London, England, 1987.

[4] R. Bajcsy, F. Solina, and A. Gupta. Segmentation versus object representation – are they separable? In R. Jain and A. Jain, editors, *Analysis and Interpretation of Range Images*, Springer-Verlag, New York, 1990.

[5] A. H. Barr. Superquadrics and angle-preserving transformations. *IEEE Computer Graphics and Applications*, 1(1):11–23, 1981. Reprinted in tutorial.

[6] A. H. Barr. Global and local deformations of solid primitives. *Computer Graphics*, 18(3):21–30, 1984. Reprinted in tutorial.

[7] R. Bergevin and M. D. Levine. Extraction of line drawing features for object recognition. In *Proceedings of the 10th International Conference on Pattern Recognition*, pages 496–501, Atlantic City, NJ, June 1990.

[8] P. J. Besl. *Surfaces in Range Image Understanding*. Springer-Verlag, 1988.

[9] I. Biederman. Human image understanding: recent research and theory. *Computer Vision, Graphics, and Image Processing*, 32:29–73, 1985.

[10] T. O. Binford. Visual perception by computer. *Proceedings of the IEEE Conference on Systems and Control*, Miami, 1971.

[11] M. Brady, J. Ponce, A. Yuille, and H. Asada. Describing surfaces. *Computer Vision, Graphics and Image Processing*, 32(1):1–28, 1985.

[12] J. F. Canny. *The complexity of robot motion planning*. PhD thesis, Massachusetts Institute of Technology, 1987.

[13] F. P. Ferrie, J. Lagarde, and P. Whaite. Darboux Frames, Snakes, and Super-Quadrics: Geometry from the Bottom Up. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-15(8):771–784, 1993.

[14] P. Fua and A. J. Hanson. Objective functions for feature discrimination. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1596–1602, Detroit, MI, August 1989. Morgan Kaufman.

[15] A. Gupta and R. Bajcsy. Surface and volumetric segmentation of range images using biquadrics and superquadrics. In *Proceedings International Conference on Pattern Recognition*, The Hague, pp. 158–162, 1992.

[16] T. Choi, H. Delingette, M. DeLuise, Y. Hsin, M. Hebert, and K. Ikeuchi. A perception and manipulation system for collecting rock samples. In *Proceedings NASA Symposium on Space Operations, Applications and Research*, Albuquerque, NM, 1990.

[17] D. D. Hoffman and W. A. Richards. Parts of recognition. *Cognition*, 18:65–96, 1985.

[18] C. Koch and S. Ullman. Selecting one among many: A simple network implementing shifts in selective visual attention. Technical Report A.I. Memo 770, C.B.I.P. Paper 3, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, January 1984.

[19] J. Koenderink and A. van Doorn. The shape of objects and the way contours end. *Perception*, 11:129–137, 1982.

[20] D. J. Kriegman and J. Ponce. On recognizing and positioning curved 3D objects from image contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990.

[21] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[22] A. Leonardis, A. Gupta, and R. Bajcsy. Segmentation as the search for the best description of the image in terms of primitives. In *Proceedings of the Third International Conference on Computer Vision*, pages 121–125, Osaka, Japan, 1990. IEEE.

[23] A. Leonardis and R. Bajcsy. Finding parametric curves in an image. In G. Sandini, editor, *Proceedings of The Second European Conference on Computer Vision—ECCV-92*. Springer-Verlag, 1992. LNCS-Series Vol. 588.

[24] A. Leonardis. Gradnja in modeliranje parametričnih struktur v slikah. *Elektrotehniški vestnik* 58(3-4):133–141, 1991.

[25] A. Leonardis. *Image Analysis Using Parametric Models: Model-Recovery and Model-Selection Paradigm*. A Dissertation in Computer Science. University of Ljubljana, Faculty of Electrical Engineering and Computer Science. May 1993.

[26] D. G. Lowe. *Perceptual organization and visual recognition*. Boston, Kluwer, 1985.

[27] D. Marr. *Vision*. San Francisco, Freeman, 1982.

[28] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Transactions on Pattern Analysis and machine Intelligence*, PAMI 15(6):580–591, 1993

[29] R. Mohan. *Perceptual Organization for Computer Vision*. PhD thesis, University of Southern California, Los Angeles, California, August 1989. Technical Report IRIS 254.

[30] A. P. Pentland. Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28(2):293–331, 1986.

[31] A. P. Pentland. Recognition by parts. In *Proceedings First International Conference on Computer Vision*, London, pp. 612–620, 1987.

[32] A. P. Pentland. Part segmentation for object recognition. *Neural Computation*, 1:82–91, 1989.

[33] A. P. Pentland. Automatic extraction of deformable part models. *International Journal of Computer Vision*, 4:107–126, 1990.

[34] A. P. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and machine Intelligence*, PAMI 13(7):715–729, 1991.

[35] N. S. Raja and A. K. Jain. Recognizing geons from superquadrics fitted to range data. *Image and Vision Computing Journal*, 10(3):179–190.

[36] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):131–147, 1990.

[37] F. Solina and A. Leonardis. Selective scene modeling. In *Proceedings of the 11th International Conference on Pattern Recognition*, pages A:87–90, The Hague, The Netherlands, September 1992. IAPR, IEEE Computer Society Press.

[38] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI 13(7):703–714, 1991.

[39] R. Volpe and P. Khosla. Manipulator control with superquadric artificial potential functions: Theory and experiments. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(6):1423–1436, 1990.

Aleš Leonardis received the Dipl. Ing. degree in 1985 and the M.S. degree in 1988 both in electrical engineering from the University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia. From 1988 to 1991 he was a Research Assistant in the General Robotics and Active Sensory Perception Laboratory at the University of Pennsylvania. He received the Doctorat degree in computer science in 1993 from the University of Ljubljana, Slovenia. Currently, he is an assistant in The Department of Computer Science at the Faculty of Electrical Engineering and Computer Science, University of Ljubljana. His research interests include computer vision, shape representation, multisensor fusion, and robotics.

Franc Solina received the Dipl. Ing. degree in 1978 and the M.S. degree in 1982 both in Electrical Engineering from the University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia. He spent five years in the General Robotics and Active Sensory Perception Laboratory at the University of Pennsylvania where he received a Ph.D. degree in computer science in 1987. He is now an associate professor of computer science and director of Computer Vision Laboratory at the Faculty of Electrical Engineering and Computer Science, University of Ljubljana. His main research interests are model based computer vision, visualization and multimedia.

Alenka Macerl is completing her undergraduate study of computer science at the Faculty of Electrical Engineering and Computer Science, University of Ljubljana. She works now as a software engineer for Hermes-Soft, Ljubljana, Slovenia.